Generative Model for the Creation of Musical Emotion, Meaning, and Form

David Birchfield
Arts, Media, and Engineering Program
Institute for Studies in the Arts
Arizona State University
480-965-3155
dbirchfield@asu.edu

ABSTRACT

The automated creation of perceptible and compelling large-scale forms and hierarchical structures that unfold over time is a nontrivial challenge for generative models of multimedia content. Nonetheless, this is an important goal for multimedia authors and artists who work in time-dependent mediums. This paper and associated demonstration materials present a generative model for the automated composition of music.

The model draws on theories of emotion and meaning in music, and relies on research in cognition and perception to ensure that the generated music will be communicative and intelligible to listeners. The model employs a coevolutionary genetic algorithm that is comprised of a population of musical components. The evolutionary process yields musical compositions which are realized as digital audio, a live performance work, and a musical score in conventional notation. These works exhibit musical features which are in accordance with aesthetic and compositional goals described in the paper.

Categories and Subject Descriptors

I.6.5[**Simulation and Modeling]**: Model Development – *Modeling methodologies*.

General Terms

Algorithms, Performance, Design, Human Factors, Theory.

Keywords

Music, Generative System, Generative Model, Generative Arts, Arts, Composition, Genetic Algorithm, Music Cognition, Perception, Multimedia, Digital Audio, Music Theory

1. INTRODUCTION

One of the challenges for generative models for the creation of time-base multimedia content is to design a system that is capable of generating outputs that contain intelligible structures and can sustain large-scale forms. A sufficient degree of compositional complexity is essential to achieve this goal, and the generated media must be intelligible to the end user in order to produce moments of tension and release and thereby convey emotion. As a traditional musician and composer I try to balance these needs when writing individual pieces of music in order to explore my compositional interests and satisfy my desire to communicate with an audience. This struggle has lead to research and experimentation in an effort to formalize this process in a generative model for music composition.

This paper describes the design and implementation of a generative model for the automated composition of music. The model is based on models of emotion and meaning in music and employs a coevolutionary genetic algorithm to generate music which is comprised of hierarchical musical structures that are navigable and discernable to listeners. The model relies on the creation of independent musical feature trajectories to produce large-scale forms. The selection of these musical features is based on empirical and theoretical research regarding music perception and cognition to ensure that the musical results are intelligible to listeners.

In this paper I will first describe some existing approaches to generative processes for the creation of music. Then I will describe two theoretical models that provide a framework for the composition of music which can convey emotion and is intelligible to listeners. I will outline the musical and compositional goals of the project, and I will describe the design and implementation of a coevolutionary genetic algorithm for the evolution of musical components and the realization of musical works. Finally, I will describe and evaluate the musical results of this work, report feedback from other listeners, and illuminate points for improving the model.

2. EXISTING MODELS

Much of the work in developing generative models for the automatic creation of music has taken two approaches. In the first, the author examines existing musical styles and attempts to understand the essential elements of the harmonies, rhythms, and textures that define the music. Next a generative algorithm is implemented which produces music that is unique at the surface level, but adheres to these stylistic principles.

David Cope has led much of this research with successful models of Bach that generate compelling musical results [1][2]. This work is able to capture the formal and grammatical features of a given style, but is bound to that particular style and cannot extrapolate to generate new grammars for yet unheard or unimagined musics.

A second approach is to take an existing generative process such as cellular automata, Lindenmeyer Systems, or games and "auralize" the data. Much of this work, including that of composer Gary Lee Nelson, has also lead to novel results [8]. However, by definition, this method is bound to the structures and patterns of the original process which, though fascinating, may not contain patterns which are truly musically compelling and communicative.

In the research described in this paper, I have endeavored to design and implement a generative system that is capable of producing musically compelling works which exhibit formal and structural coherence that is stylistically unique and discernable to the listener. I have drawn on two theories of emotion and composition to design the generative system.

2.1 Models of Emotion and Meaning

In his seminal work, *Emotion and Meaning in Music*, Leonard Meyer proposes a model of emotion in music that is based on fulfillment and inhibition of a listener's expectations [7]. He explains that "one musical event (be it a tone, a phrase, or a whole section) has meaning because it points to and makes us expect another musical event." Expectations can be used to generate emotions such as contentment in a piece of music, "because of our belief in the presence of control and in the nature of the resolution, [resolutions] prove most pleasurable." Alternatively, if a listener is unable to navigate a piece of music, or if expectations are always thwarted, "Ignorance and its concomitant feelings of impotence breed apprehension and anxiety, even in music."

Meyer posits that there are two aspects to generating emotionally charged moments in music. First, as a prerequisite, the listener must be capable of understanding the musical structures in either a conscious or unconscious manner. This understanding is either acquired through exposure to a musical style in general or a piece of music in particular. Secondly, once a sense of expectation is created, feelings of fulfillment and stability can be created by resolving the expectation. Suspense, and even anxiety, can be sustained by suspending the resolution.

2.2 Compositional System vs. Listening Experience

Meyer's theory is based on a typical listener's experience of listening to common practice tonal music of composers such as Bach, Mozart, and Beethoven. The theory can be extended to include styles such as Jazz, Blues, and Rock given that the tonal pitch system is prevalent in these musical styles as well. However, there can be a disconnect between the compositional aims of a piece of music and a listener's ability to perceive structure when other, potentially novel, style systems are used.

Fred Lerdahl describes this divide in his article, "Cognitive Constraints on Compositional Systems [5]." The article focuses on the highly sophisticated compositional processes of serial composers which often yield opaque musical results that are distanced from the means of production. He writes, "Boulez's *Le Marteau sans Maitre* was widely hailed as a masterpiece of postwar serialism. Yet nobody could figure out, much less hear, how the piece was serial."

Lerdahl proposes that a listener's ability to cognize the musical surface must be taken into consideration in the composition of music to avoid "a huge gap between compositional system and cognized result." Compositional systems that ignore this paradigm will fail to communicate with listeners.

3. MUSICAL GOALS

In designing this generative algorithm, I have drawn on these two important theories in order to generate music that is emotionally communicative and capable of conveying a sense of mood and tone, but not based on an existing stylistic model. The music must contain structures which are intelligible to the ear, and these structures must be both stable and flexible to allow for recognition of normative and distorted versions. The introduction of distorted and manipulated versions of a normative structure establishes a musical expectation for resolution. As described in Meyer's theory of emotion, fulfillment of this expectation leads to a release, while tension can be sustained by prolonging the sense of deviation. The precise configurations and modes of tension and release do not need to be specified, so long as they emerge in some discernable fashion from the generative process.

I have defined four broad musical goals that I hope the system can achieve. They are open in that they can accommodate many different surface-level characteristics, but restrictive to the extent that they define the basis of music that I hope the generative system will produce.

1. I want to create normative structures that can be readily manipulated and distorted to create musical forms which are discernible to the listener. These structures are in the frequency domain, but also rhythmic, metric, and timbral.

2. I hope to generate large, slowly evolving forms which are balanced by local variety.

3. Having chosen a set of musical features to explore in a given piece, I strive to include the extremes of each feature in the same piece.

4. I hope to generate musical processes which unfold over time independently for each musical feature. The surface of the music is the composite resulting from the interaction of these otherwise independent processes.

I evaluate the success of the generative algorithm according to the musical outputs and their adherence to the goals described above. The generated music and evaluations are presented in sections 6 and 7.

4. THE GENERATIVE MODEL

This generative model for the automated composition of music utilizes a coevolutionary genetic algorithm. The algorithm is hierarchically organized and is comprised of a population of musical components including Notes, Gestures, Phrases, Sections, and Meta Sections.

I chose to implement a genetic algorithm for this work because the computing paradigm is sufficiently flexible and open-ended to allow for manipulation and enhancement to suit the musical purpose. Nonetheless, it is also possible to leverage features of the genetic algorithm that are well suited for organizing music.

1. An unlimited genetic feature space can be defined for each member of the population.

2. Each genetic feature evolves independently to yield virtually unlimited combinations.

3. Members of the population will mate with one another to produce children which are a hybrid of the two parents. For example, two musical Sections can mate to produce a third Section which will exhibit musical features such as frequency and amplitude which are drawn from both parents. In this manner, the genetic algorithm affords the possibility for repetition and distortion in time. This fact allows for the cross-pollination of salient musical features at various hierarchical levels.

4. A genetic algorithm is a complex nonlinear system and one cannot predict the outcome of the evolutionary process without actually running the simulation. Genetic mutation and the infinite variety of combinatorial possibilities, allow the algorithm to generate novel and unpredictable results. I want to build a generative system that is capable of producing musical results that are not explicitly built into the system and could not have been anticipated. New forms and musical ideas can emerge from the system which adhere to the goals outlined above, but are not 'composed' in a traditional sense.

4.1 Feature Trajectories Drive Form

In my work as a composer, I often sketch curves and lines that plot the trajectories of how musical features will develop throughout a given piece of music. For example, I might imagine a piece that exhibits a timbral evolution moving from bright to dark and a density trajectory from dense to sparse. Simultaneously, the amplitude level might rise from low to high, and return to a moderate level toward the end. I have found that this sort of design and realization of multiple, independent musical feature trajectories allows for complex musical structures to emerge. I have worked to design this capability into the generative model. Such structures can support large-scale forms of durations ranging from ten to thirty minutes.

To provide the capability of generating such independent feature trajectories, I have expanded the scope of the genetic feature to include not only a primary value, but also a vector of attributes that define contours for child components of the hierarchy. For example, a Phrase in the evolutionary system can define that the first of seven child Gestures should begin at a low frequency, move toward a high frequency by the fifth Gesture, and by the seventh it should return to a medium frequency. These basic feature contours are defined by a start value, middle value, middle value location, and an end value that serve as anchor points for the skeleton of the curve. The bounds of these feature trajectories are limited by a feature bandwidth attribute. Additionally, the bandwidths of child components can be shaped along trajectories that are specified by bandwidth start and end anchor points.

The fitness function of the components is defined by the bandwidth and trajectory attributes of the parent. As a consequence, fitness in the system is a dynamic, ever changing standard.

Artificial intelligence is employed by individual components in the system to self-organize themselves along the skeleton trajectory which is defined by the anchor points of the parent. This intelligence allows for local, context-aware interactions between child components. This interaction leads to an infinite variety of possible feature contours in the generated structures; trajectories which are more varied and complex than if simply defined by three anchor points.

The additional of this expanded set of feature attributes and local interactions does introduce an extra degree of conceptual and computational complexity. Nonetheless, this methodology yields diverse and musically compelling feature trajectories, and can be broadly applied to other domains of time-based multimedia creation.

5. MUSICAL FEATURE SET

The genetic algorithm described above provides a means of generating a population of musical components that contain independent feature contours. Given this powerful tool, the next challenge is to identify those musical features which will lead to satisfying compositional results and are salient to listeners. Features such as frequency, amplitude, and duration are easily identified and defined, but in order for sophisticated, hierarchical musical constructs to emerge from the system, many more features must be utilized. Furthermore, these features must traverse a trajectory from normative to distorted in order to be useful in serving the system's model of emotion and meaning.

An example feature of the system is *Pitch Clarity*. At its highest level, this feature will define sounds which have a clear core frequency. A sine tone has absolute pitch clarity. This is the normative, readily understandable end of the spectrum. At the other end is a highly noisy sound such as pink or white noise. This example is highly distorted as no frequency information is conveyed by noise.

5.1 Frequency and Harmony

Of course, many musical features are interdependent on one another. Frequency is a feature of individual Notes, but a collection of Notes will create harmonies that are can be classified as consonant or dissonant, normative or distorted. The generative model implements a number of musical features which are collectively used to create harmonies. The organization of elements in the frequency domain is based on research by cognitive psychologists such as Caroline Krummhansl [4] and Ernst Terhardt [10]. It also employs a model of dissonance and consonance which draws on theories of harmonic roughness proposed by Richard Parncutt [9].

5.2 Rhythm and Meter

Like harmony, rhythm and meter are high-level musical structures that depend on a collection of sub-features for their creation. As an example of how such structures can emerge from the system I will examine the rhythmic features of the system in more detail.

5.2.1 Meter

The design and implementation of the feature set for generating meter is based on theories of metric construction and perception [6][3]. This research finds that the establishment of a stable metric grid depends on the creation of regular, repeated primary beats.

In the genetic algorithm, three features define the details of a Gesture's underlying primary pulse. These features are the *Number of Rhythmic Primary Pulses*, *Primary Pulse Omission Probability* and the *Focal Primary Pulse*. Figure 5.1 illustrates a Gesture with seven equally spaced primary pulses. This regularly pulsing establishes a metric grid without interruption that is a normative, stable structure analogous to a consonant harmony or a note with absolute pitch clarity.

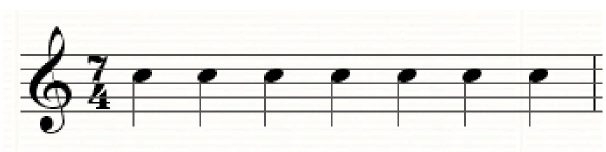

Figure 5.1

In figure 5.2, the *Pulse Omission Probability* has been introduced and the stability of the metric grid is slightly degraded.

Figure 5.2

In Figure 5.3, the *Focal Primary Pulse* has been defined and is indicated by the accent on the fifth beat.

Figure 5.3

5.2.2 Rhythm

Once a regular pulse has been established as described above, those pulses can be subdivided by smaller rhythmic values. In designing this subdivision scheme I wanted to generate rhythmic Gestures that would have fluid accelerations and decelerations around the focal pulse. I also wanted to allow for a wide range of possible subdivisions. Finally, I wanted to be able to generate both regular subdivisions that reinforce the underlying metric grid, and also more dispersed subdivisions that would distort the sense of a regular pulsing. The number of attacks within a subdivision can range from 0 – 7 where 0 is the normative, least distorted value, and seven is the greatest distortion of the primary pulse. This model draws on David Epstein's research and theory of rhythmic perception and performance which finds that rhythms which are related by simple ratios are more easily perceived and understood than those related by complex ratios [3].

Figures 5.4, 5.5, and 5.6 show various subdivisions of the metric grid which could be generated by the system. Each example illustrates an increasing degree of distortion of the underlying seven primary pulses.

Figure 5.4

Figure 5.5

Figure 5.6

Finally, in Figure 5.7, a *Subdivision Omission Probability* feature has been introduced. This example illustrates a highly distorted realization of the underlying normative metric grid.

Figure 5.7

6. MUSICAL RESULTS

6.1 Evolution is Independent from Realization

Thus far, I have described the components that comprise the musical hierarchy in abstract terms. Each member of the population has a set of genetic features which define its musical qualities, but until a given component is realized as an actual musical event, these Phrases, Gestures, and Notes are merely abstract data structures. The evolution of musical structures is independent of their ultimate realization. The final step in the generative process is the realization of the musical population as a piece of music. In this section I will broadly discuss issues pertaining to this realization process.

6.2 Realization Methods and Biases

The musical realization of a population of notes is a separate process from their evolution. Except in the broadest terms, the Notes are not evolved with a particular instrumentation or type of computer processing in mind. When a set of Notes are to be realized, they are passed along to a software agent which has a definition of the capabilities of the relevant instruments or computer processing means. That agent then examines the feature set of each Note, and realizes the given Note with the best possible instrument or computer process to best express those features.

To date I have used four methods to realize the musical components. All methods are used to realize populations which are evolved in an identical fashion, but each realization highlights certain aspects of the generated musical structures.

Firstly, all Notes of the system have been realized as Triangle Tone Oscillators. This realization method is extremely precise with regard to rhythms, amplitudes, and frequencies. Consequently, the harmonic and metric aspects of the resultant music is extremely clear. However, timbral information is totally absent in these realizations.

Secondly, the Notes of the system have been realized as filtered and unfiltered soundfiles that are drawn from an indexed database of digital audio files. In this realization, timing, amplitude, and pitch clarity data is very accurate, but often the frequency information can be distorted or lost all together. This method allows local rhythmic gestures and timbral trajectories to become most present.

Thirdly, the Notes of the system have been realized in a hybrid fashion such that some are realized as filtered soundfiles by the computer in a fashion similar to that described above, and the others are realized by a live musician who sightreads and interprets the graphical score generated by the software GUI. In this case some rhythms and amplitudes are accurately realized, but many are distorted and altered by the live performer. Here, phrase structures and motivic repetitions are especially brought to the listener's attention.

Finally, the system has been converted in to a musical score in standard notation that is given to an ensemble of musicians for live performance. In this version, many concessions must be made to make the music performable by human players on standard instruments. Consequently, many details of frequency, amplitude, and rhythm are approximated or altered. Nonetheless, phrase densities, tempo fluctuations, and large-scale forms are clarified and accentuated in performance.

This method of using multiple realization methods has been very informative. Although to a certain extent, these pieces should all sound relatively uniform given that they make use of the same set of musical features in an attempt to generate musical forms and structures, in practice they vary greatly. As described above, each realization method accents different musical aspects of the generative process, and the sense of style in these pieces is more dependent on the realization method than the abstract organization of the musical structures. All are equally successful at conveying emotion and meaning through the generation of discernable musical structures, but they succeed along different dimensions.

7. CONCLUSIONS

7.1 Evaluation of Generated Works

The true measure of the success of the generative algorithm is the quality of the music it generates. 'Quality' is defined according to adherence to the points described in section 3. Over the course of working on this project I have been encouraged by the steady improvement in the quality of the resulting music. The first musical outputs were undifferentiated from one another, and exhibited flat contours at the formal and local levels. As the feature dimensions were enhanced, more varied trajectories began to emerge. As the feature list was expanded, the generated pieces grew in sophistication and clarity. This improvement in the musical results is evidence of an increased understanding of the critical features of the music I hoped to generate and my ability to quantify their influence.

As the project progressed, I also felt that as a listener I was increasingly able to navigate the musical structures which emerged and could more clearly recognize my own musical biases in the pieces. This ability to grow more conditioned to the musical style of the genetic algorithm provides further evidence of the increasing sophistication of the generated music. This progress marks the achievement of one of the primary goals of generating normative musical structures that can be discerned by a listener.

Another musical goal of the project was to generate music which exhibits large-scale, slowly evolving forms and addresses the passage of time on multiple levels. I am pleased that the genetic algorithm produces pieces that exemplify such forms.

A final musical goal was to generate music that would reach the extremes of each musical feature in the system. Though in some cases I felt that this goal was achieved, this is an area for future improvement in the algorithm. In particular, the rhythmic language of the generated music is not as varied as I would ultimately hope for and I expect that future work on the generative system will improve this area.

7.2 Listener Feedback

Undoubtedly the results of the generative process should also be evaluated according to feedback from colleagues and other listeners.

In general, listeners report that the generated music provides a challenging but rewarding listening experience that requires careful attention and concentration. Though they may not necessarily be able to articulate the influence of exact musical mechanisms, they explain that the music does establish a cohesive whole with local variety and interest that they can intuitively navigate.

Some expert listeners articulate that the generated music exhibits a focused intensity which is primarily driven by shifting rhythms, gestures, and phrasing. They also describe that, while the music often contains novel and exciting moments, these moments are sometimes tempered by more idle sections that seem to stall the musical momentum

I hope that further improvements discussed in section 7.3 will allow the generative model to produce results that are more directly communicative. Nonetheless, these responses are satisfying and encouraging for future directions with the model.

7.3 Points for Improvement

Despite the fact that the musical pieces grew in sophistication over the course of this work, there is more work to be done. As other listeners describe, I too find that there are often moments in the generated pieces where the drive of the music seems to temporarily stall. I specified that formally I want the music to unfold slowly over time, but I anticipate that further improvements to the algorithm can prevent these static moments.

In some cases, the musical outputs of the system could benefit from a reduction in the number of active musical features. The introduction of nearly seventy-five, independently evolving musical features increases the variety and potential for novel musical structures to emerge. However, this degree of complexity can sometimes lead to an over abundance of information. Reducing the sheer number of features might maximize the power and clarity of those that remain, and thus some of the musical forms, which are currently too fluid, might be clarified.

Finally, the realm of musical timbre is underdeveloped in the generative system. The algorithm presently includes a Pitch Clarity feature as described in section 5, but in future work I hope to more fully develop the timbral aspects of the generated music.

8. REFERENCES

[1] Cope, D., Virtual Music: Computer Synthesis of Musical Style. The MIT Press, Cambridge, MA, 2001.

[2] Cope, D., Computers and Musical Style. A-R Editions, Madison, MI, 1991.

[3] Epstein, D., Shaping Time. Schirmer Books, New York, 1995.

[4] Krumhansl, C., Cognitive Foundations of Musical Pitch. Oxford University Press, Oxford, 1990.

[5] Lerdahl, F. Cognitive Constraints on Compositional Systems. Contemporary Music Review, vol. 6 (1992).

[6] Lerdahl, F., and Jackendoff, R., A Generative Theory of Tonal Music. The MIT Press, Cambridge, MA, 1983.

[7] Meyer, L.B., Emotion and Meaning in Music. The University of Chicago Press, London, 1956.

[8] Nelson, G.L., Fractal Mountains. Computer Music Currents, vol. 10 (1992).

[9] Parncutt, R. Harmony: A Psychoacoustical Approach. Springer-Verlag, Berlin, 1989.

[10] Terhardt, E., Pitch, Consonance and Harmony. Journal of the Acoustical Society of America, vol. 55, No. 5 (1974), 1061-1069.